EXECUTIVE SUMMARY

TABLE OF CONTENT.

1. Project Introduction
2. Loading of Dataset
3. Data Exploration
4. Dealing with DateTime Features
5. Unique Values in Columns
6. Aggregation with Groupby
7. Data Visualization.

8) Conclusion.

(1) Projection Introduction

Company XYZ owns a supermarket chain across the country. Each major branch located in 3 cities across the country recorded sales information for 3 months, to help the company understand sales trends and determine its growth, as the rise of supermarkets competition is seen to increase. The data folder contains datasets from three different branches; Lagos, Abuja and Port Harcourt. Each data file from the branches contains the same attribute information.

We going to analyze the different branches, comparing their columns to better understand the datasets and in making a better marketing strategy and increase the supermarket profit.

Some columns on the data set include

 Invoice ID: Customer Identification number

Branch: Supermarket Branch across the country (A, B, C)



A - Lagos Branch

B - Abuja Branch

C - Port Harcourt Branch

 City: Supermarket Location

Customer Type:



Type of customers, Members - Returning customer with membership card, Normal - Customer without membership (could be returning, first-time or walk-in customer)

Gender: Customer Gender Information(Male or Female)



Product line: Product categorization groups - Electronic accessories, Fashion accessories, Food and beverages, Health and beauty, Home and lifestyle, Sports and travel



Unit Price: Price of each product in Naira



Quantity: Number of products purchased by customer



Tax: 5% tax fee for customer buying



Total: Total price including tax



Date: Date of purchase (Supermarket Record available from January 2019 to March 2019)



Time: Purchase time (Supermarket Hours - 10am to 9pm)



Payment: Payment used by customer for purchase (3 methods are available – Cash, Card and Epay)



COGS: Cost of goods sold



Gross margin percentage: Gross margin percentage



Gross income: Gross income



Rating: Customer Satisfaction rating on their overall shopping experience (On a scale of 1 to 10)



2) Loading of Dataset

Loaded 3 different csv dataset or files. Each of the dataset are made up 3 Branches namely Branch A = Lagos branch, B = Abuja branch, C = Port Harcourt. The .csv files was passed in List using .glob(). The 3 files was joined together as one dataset using pd.concat and the dataset was read using pd.read\_csv format.

3) Data Exploration:

Pandas was imported to convert dataset to Dataframe while numpy was also used to Numerical computation. seaborn and matplotlib was also use for plots. i used the .head() to preview the dataset, the isnull to check for missing values. i also used the .info() to see the summary of the data frame. The data contains 1000 rows and 17 columns.

key things i noted was that the highest rating is 10 and lowest is 4.0. while the maximum gross income is 17874. we also noticed that the highest unit price as 35985 and minimum as 3661.

4) Dealing with DateTime Feature

Here i dealt with DateTime features. I converted Date to date and time into the date format using pd.to\_datetime. i also extracted the hour, Day, Month and Year from our Date and Time Column

5) Unique Values in Columns

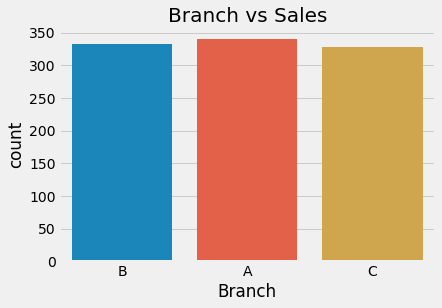
I was able to get Unique Value in Column iteration and i got a categorical column. i now used .unique().tolist() function to get the unique value from each categorical column. Also generated the count figure of the categorical values using the value\_counts() method.

6) Aggregation with Groupby

Here firstly I did a groupby object with City column using and aggregate function was used to get the sum and mean. Also determined the gross income across all cities and it was gotten that Port Harcourt has the highest. Did a Unit price and Quantity analysis, Here Lagos has the Quantity and Unit Price across all Cities.

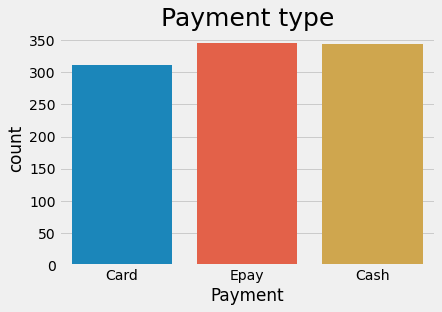
7) Data Visualization

I used charts and plots to create visualization to answer some questions like Cites with the highest Sles, Most used payment methods across cities and branches then many more.



Insight

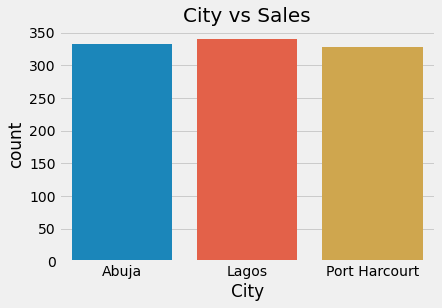
Branch A did the highest sales as seen from the plot followed by branch B



Insight

1) The most and preferred payment type is Epay followed by Cash.

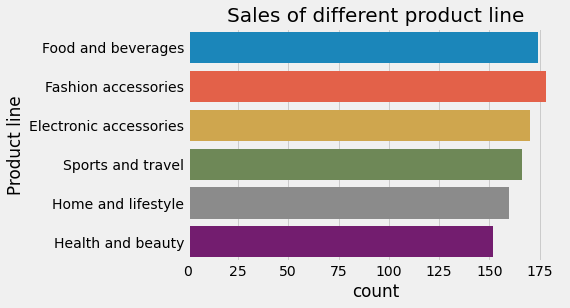
2) The least used payment method is Card.



Insights

1) Lagos is the city with the highest sales with Abuja following.

2) Then Port Harcourt has the least.

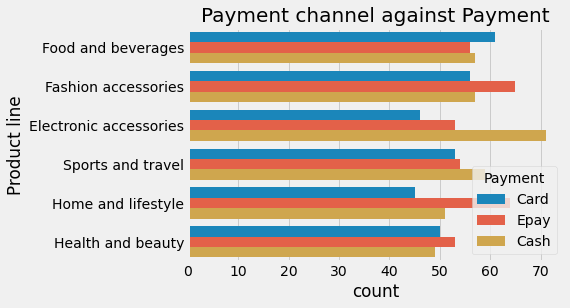


Insights

1) The highest product line as seen from the plot is Fashion accessories

2) The lowest product line is Health and beauty

3) Immediately after the first is Food and beverages which is ranking second in the plot.



Insights

1) As seen from the plot the most used payment channel in Food and Beverages is Card and the less used is Epay

2) From Fashion accessories the most used payment channel is Epay and less used is Card

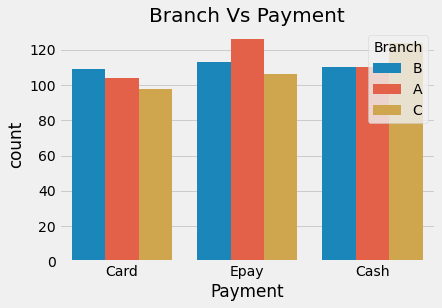
3) From Electronics accessories the most used payment channel is Cash and less used is Card

4) In Sports and travel the most used payment method is Cash and the least used is Card

5) Looking at the plot you will notice in Home and Life style that the most used payment method is Epay and least used is Card

6) Health and beauty Product line has more of Epay used as payment method than any and Cash as the least payment method.

Generally, we can clearly see that Epay as the most used across all Payment method and Card as the least.



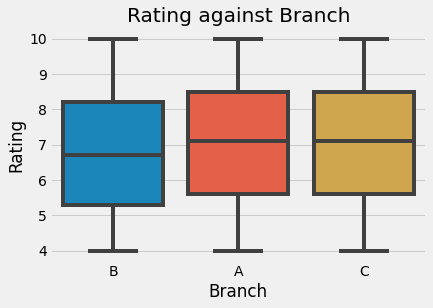
Insights

1) Branch B uses more of Epay as a payment method compared to Card and Cash

2) Branch A uses more Epay than than any other payment method. We have Cash as the second most prefered method in Branch A

3) Branch C uses Cash more than any other payment method and Epay is the second preferred bfore Card.

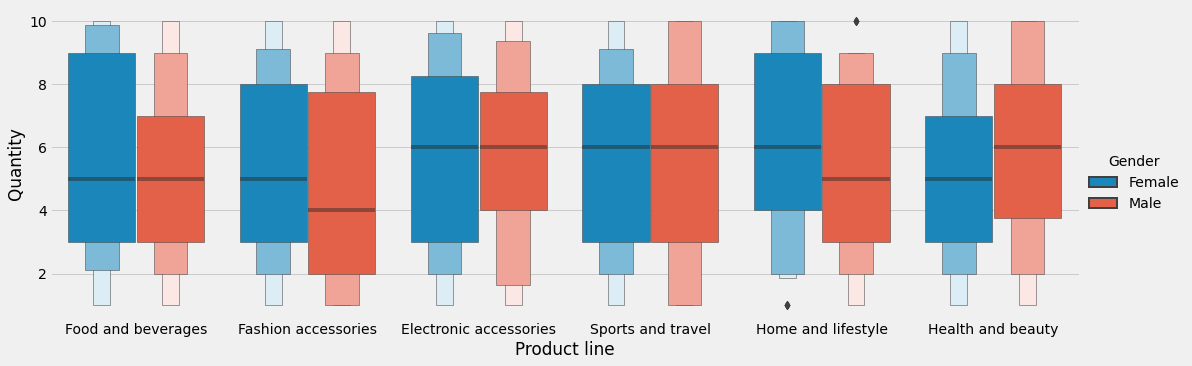
Epay generally remain the most used considering all payment method.



Insight

The Branch A and C are the most highly rated Branches while B is the least.

Product line Vs Quantity



Insights

1) Generally from the catplot we can see that we have males buying more quantity of Food and Beverages, Fashion accessories,

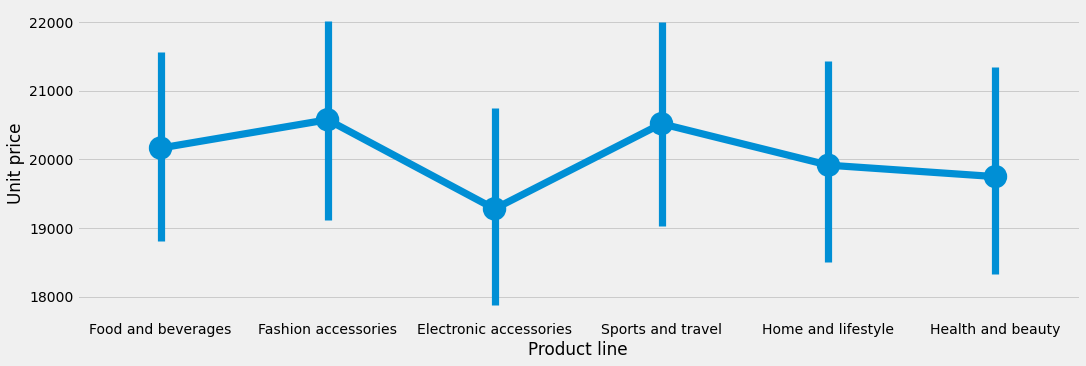
Electronic accessories and Home and Lifestyle

2) We have male and female buying equal quantity of products in Sports and travel.

3) Finally Female bought the highest quantity of products in only Health and beauty.

Generally considering the catplot we see male make more purchase across all product line than Female.

Product line Vs Unit Price

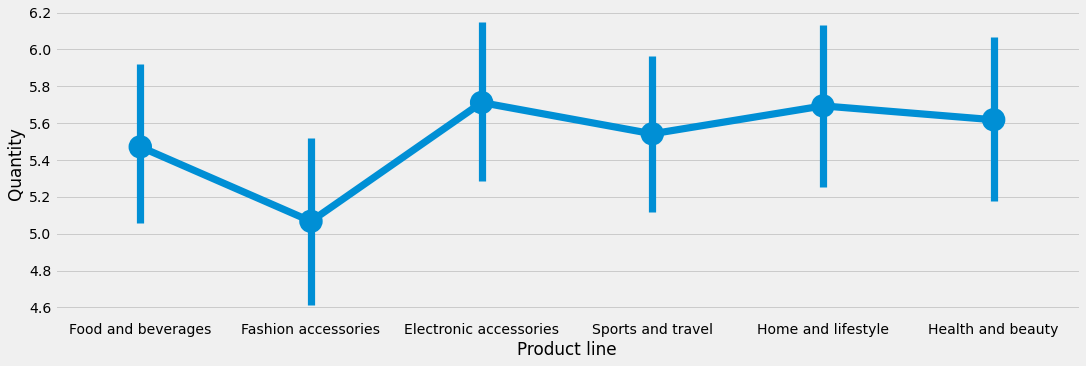


Insights

1) Fashion accessories, Sports and travel has the highest unit sales across all Product line

2) Electronic accessories have the lowest unit price across all Product line.

Product Line Vs Quality

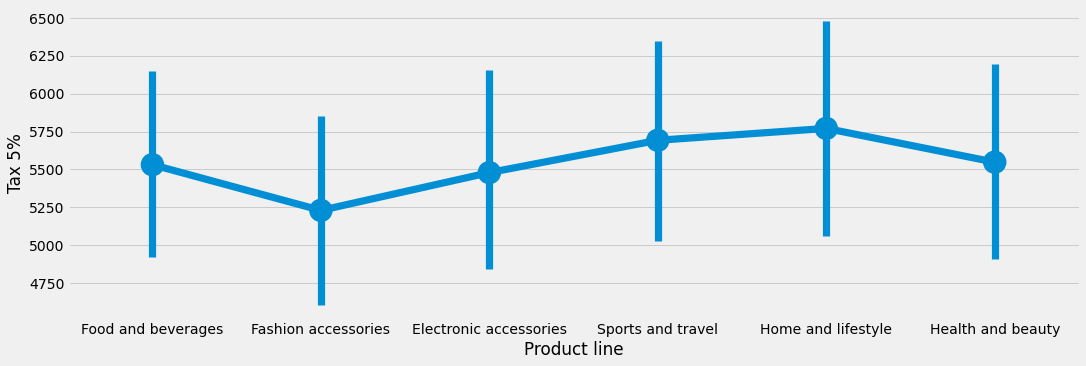


Insights

1) Electronic accessories, Homes and lifestyle and Health and Beauty product line have the Highest Quantity sales across all respective product line

2) Fashion accessories have the lowest Quantity sold across all Product line.

Product Line Vs Tax



Insights

1) Home and style paid Highest tax across all product line.

2) Fashion paid the least tax across all product line

Conclusion.

In conclusion we can clearly deduce these.

1) We can agree that since Epay the most used means of payment it should be always available so there will be no restriction in payment for products.

2) They company should invest more on Fashion and accessory and Food and Beverages since they are the most sold product line.

3) They should Invest more in Port Harcourt because that’s the location the make the most income.

4) The Company should consider buying more of ladies’ products because the make the most purchase than Male except for the Product Line “Health and Beauty”

5) They should also Invest in Lagos since the had the largest sales across all Cities.

6) Branch A made the largest sales so they Company should channel more resources to branch A.